3168–3173 Nucleic Acids Research, 1995, Vol. 23, No. 16

The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1

Rein Aasland*and A. Francis Stewart

Gene Expression Programme, EMBL, Meyerhofstraße 1, D-69117 Heidelberg, Germany

Received May 30, 1995; Revised and Accepted July 19, 1995

ABSTRACT

The chromo domain was originally identified as a protein sequence motif common to the *Drosophila* chromatin proteins, Polycomb (Pc) and Heterochromatin protein 1 [HP1; Paro and Hogness (1991) *Proc. Natl. Acad. Sci.* USA, 88, 263–267; Paro (1990) *Trends Genet.*, 6, 416–421]. Here we describe a second chromo domain-like motif in HP1. Subsequent refined searches identified further examples of this chromo domain variant which all occur in proteins that also have an N-terminally located chromo domain. Due to its relatedness to the chromo domain, and its occurrence in proteins that also have a classical chromo domain, we call the variant the 'chromo shadow domain'. Chromo domain-containing proteins can therefore be divided into two classes depending on the presence, for example in HP1, or absence, for example in Pc, of the chromo shadow domain. We have also found examples of proteins which have two classical chromo domains. The *Schizosaccharomyces pombe* SWI6 protein, involved in repression of the silent mating-type loci, is a member of the chromo shadow group. The similar modular architecture of SpSWI6, HP1 and HP1-like proteins supports the model that the specificity of action of chromatin proteins is generated by combinations of protein modules.

INTRODUCTION

Genetic analyses of *Drosophila* and yeasts have identified several classes of proteins involved in the regulation and maintenance of chromatin (1,2). Most of these proteins do not appear to be DNA binding proteins that function as conventional transcription factors. Their mechanisms of action remain the subject of speculation, and a model has emerged which invokes combinatorial specificity and the formation of multi-component complexes (3,4). Key to the current favourable regard for this model was the observation that Polycomb (Pc) shares a protein motif with a cytologically identified heterochromatin protein, HP1 (5,6). This motif, termed the chromo domain, links Pc, a genetically identified repressor of *Drosophila* HOM-C expression, with a protein involved in heterochromatin-mediated repression. Furthermore, HP1 was identified as the product of the *Su(var)205* locus, thus confirming the link between repression mediated by heterochromatin and position effect variegation (PEV; 7). The importance of discrete protein motifs in chromatin regulation gained further impetus from the observation that Enhancer of zeste [E(z)] protein, encoded by a locus genetically defined to repress HOM-C expression (8) shares a protein motif, termed the SET domain, with trithorax (trx) (9). Trx is the defining member of the trithorax group of positive regulators of HOM-C expression. Thus the SET domain is shared between members of both the positive and negative regulators of HOM-C expression suggesting it is important in the regulation of HOM-C expression. Recently, we observed that a newly identified protein domain, a novel zinc finger, termed the PHD finger, is also shared by trithorax and another member of the Pc group of HOM-C repressors, Polycomblike (Pcl) (10).

Shared protein domains are central to current thinking about the mechanism of action of genetically defined chromatin activators and repressors because they suggest a basis for understanding regulation in chromatin. By first identifying these domains and then identifying their roles, it should be possible to unravel the hierarchies of interactions that determine regulation in chromatin. Here we report the identification of a new protein motif, the chromo shadow domain. Consistent with the combinatorial models of regulation in chromatin, the chromo domain occurs in several dispositions, one of which involves pairing with the chromo shadow domain.

MATERIALS AND METHODS

Sequence similarity searches and alignments

Sequence similarity searches were performed with profile analysis as implemented in the program SEARCHWISE (11; E. Birney, J.D. Thompson and T.J. Gibson, unpublished). Profiles were prepared with the program PROFILEWEIGHT version 2.0 (12; J.D. Thompson, D.G. Higgins and T.J. Gibson, unpublished) using BLOSUM45 and BLOSUM62 substitution matrices (13), optionally normalised for amino acid relative mutability. Other parameters were: branch-proportional sequence weighting; exclusion of alignment positions with >50% gaps; gap opening and gap extension penalties at existing gaps in the alignment were 5% of standard values.

Initial searches used a profile based on an alignment of well characterized chromo domains (14). Chromo domains detected by these searches were subsequently aligned with the program CLUSTAL W (15) and manually edited with the GDE alignment

* To whom correspondence should be addressed at present address: Laboratory of Biotechnology, University of Bergen, HiB, Thormöhlensgt. 55, N-5020 Bergen, Norway

Table 1. Chromo domain-containing proteins

domain name[1]	database entry[1]	Acc.No.	score[2] classic	score[2] shadow	description
Classical chromo domains.					
DrPc	sw:PC_drome	P26017	10200	4090	Polycomb
MoMOD3	sw:MOD3_MOUSE	P30658	10660	4460	M33, related to Polycomb
CeYO82	sw:YO82_CAEEL	P34618	12030	3350	Nucleolin similarity
DmHP1_A	trembl:DMHP1_1	M57574	12400	4400	Su(var)205, modifier of PEV
DvHP1_A	sw:HP1_DROVI	P29227	12300	4520	modifier of PEV
HuHP1_A	trembl:HSHP1HOM_1	L07515	12640	5120	similar to HP1
MoMOD1_A	sw:MOD1_MOUSE	P23197	12450	5320	M31, similar to HP1
MoMOD2_A	sw:MOD2_MOUSE	P23198	11800	5210	M32, similar to HP1
PcHET1_A	Epstein et al.[3]	n.a.	12710	5830	similar to HP1
PcHET2_A	Epstein et al.[3]	n.a.	12030	4440	similar to HP1
SmPAJ26	emest:SMT14583	T14583	11740	5200	EST[4]
SpSWI6_A	sw:SWI6_SCHPO	X71783	9885	4575	repressor of mating-type loci
Pf0131C	emest:PF521	T02521	8950	2900	EST[4], partial.
CeT9A58	trembl:CET09A5_12	Z36753	9185	2690	cosmid T09A5 gene 8
DmSuv3-9	trembl:DMUSVAR39_1	X80070	9025	2280	Su(var)3-9, modifier of PEV[5]
HuMG44	embl:HSMG44A	L08238	10160	3090	related to Su(var)3-9[5]
CfTENV	trembl:CFT1RTPOS_3	Z11866	11305	3130	retrotransposon
FoSKPY	trembl:FOGAGPOL_2	L34658	10200	3160	retrotransposon
MoCHD1_A	sw:CHD1_MOUSE	P40201	9490	3390	helicase-domain
MoCHD1_B	- " -	- " -	7985	4020	
CeYK9A3	emest:CEK009A3F	D27447	6100	2120	EST[4], partial, MoCHD1-like
ScYEZ4_A	sw:YEZ4_YEAST	P32657	6960	2325	YER164w, helicase-domain
ScYEZ4_B	- " -	- " -	8115	4020	
MgGRH	trembl:MGGRHA_2	M77661	9085	3060	Grashopper, retroelement
MgMAGGY	trembl:MGGAGPOLH_2	L35053	7965	2980	MAGGY, retrotransposon
Ce29H12	tremblnew:CEC29H12_2	U23169;	8970	2375	cosmid C29H12, gene 5
Chromo shadow domains					
DmHP1_B	trembl:DMHP1_1	M57574	3875	14330	modifier of PEV
DvHP1_B	sw:HP1_DROVI	P29227	8080	14770	- " -
HuHP1_B	trembl:HSHP1HOM_1	L07515	8080	15870	similar to HP1
MoMOD1_B	sw:MOD1_MOUSE	P23197	8080	15940	M31, similar to HP1
MoMOD2_B	sw:MOD2_MOUSE	P23198	4795	15730	M32, similar to HP1
PcHET1_B	Epstein et al.[3]	n.a.	5485	15220	similar to HP1
PcHET2_B	Epstein et al.[3]	n.a.	4380	11450	similar to HP1
SpSWI6_B	trembl:SPSWI6_1	X71783	2445	10510	repressor of mating-type loci

[1]The amino acid sequences used in this study were either extracted from SWISS-PROT (sw) or from TREMBL (trembl, tremblnew), the translated version of the EMBL database or from translation of ESTs (expressed sequence tags) as found in the EST-subdivision (emest) of the EMBL database. The sequence of HuMG44 was obtained by translation of the corresponding EMBL entry. The domain names have the following species abbreviations: Dm, *Drosophila melanogaster*; Dv, *Drosophila virilis* Mo, mouse; Hu: Human; Ce, *C.elegans*; Sc, *S.cerevisiae*; Sp, *S.pombe*; Cf, *Cladosporium fulvum*; Mg, *Magnaporthe grisea*; Pc, *Planococcus citri*; Pf, *Plasmodium falciparum*; Sm, *Shistostoma mansoni*. n.a.: not applicable.

[2]Scores were calculated with the program PAIRWISE using profiles based on the parts of the alignment (Fig. 2) corresponding to the classical chromo domains and the chromo shadow domains as described in methods.

[3]The sequences PcHET1 and PcHET2 were typed in from (21).

[4]ESTs derived from human genes encoding homologues of several mouse chromo domain proteins where identified. These are virtually identical to the corresponding murine sequences and therefore not included in the alignment. The following ESTs were found [database:id (accession number)]: MoMOD2 homologes: embl:HSA41C081 (Z15820), embl:HSAFIA009, (Z18797), embl:HS10835 (T64108), embl: HS82033 (T63820). The entry embl:HSC1AH092 (F02792) is also related to, but distinct from the putative HuMOD2 sequences. It was not possible to generate a reliable alignment with this sequence and it is therefore not included in this study. MoMOD3 homologue: embl:HS6418 (T11641). MoCHD-1 homologue: embl:HS8101 (T05810).

[5]The human MG44 cDNA sequence can be conceptually translated (assuming several errors in the sequence) to a protein homologous to the Su(var)3–9 protein of *Drosophila* (22). Both proteins have a chromo domain as well as a SET domain (9,22).

editor (S. Smith, Harvard University). Further exhaustive searches were performed with profiles based on the new alignments. The program PAIRWISE (E. Birney, J.D. Thompson and T.J. Gibson, unpublished) was used to check for multiple chromo domains in each sequence and to calculate the final scores (using the BLOSUM45 matrix). The profiles (see below) used for the final searches, the alignment and further information on the chromo domains is available on the World-Wide-Web (http://www.uib.no/aasland/chromo.html). A detailed description of SEARCHWISE and PAIRWISE can be found on the World-Wide-Web page (http://www.ocms.ox.ac.uk/~birney/wise/ topwise.html).

Dotplot analysis

Dotplots were generated with the program PROPLOT (12) using profiles generated with the program PROFILEWEIGHT version 2.0 (12; J.D. Thompson, D.G. Higgins and T.J. Gibson, unpublished) using alignments (position 1–72) as described above and the

residue substitution matrix BLOSUM45 (13). A window size of 13 was used for the plots, as this is close to the observed block size in chromo domains (Fig. 2). The BLOSUM45 matrix was first normalised by the amino acid relative mutabilities (T.J. Gibson and J.D. Thompson, unpublished). This correction ensures equivalently conserved positions in an alignment are equally weighted. Otherwise, for example, tryptophan at position 34 would score three times more highly than glutamic acid at position 11, even though they are equally conserved in chromo domains, hence equally important to the chromo domain signature.

Sequence space analysis

The partial sequences were removed from the multiple alignment and the resulting alignment was then subjected to 'sequence space analysis' using the programs SEQUENCE SPACE and SCATTER, a multivariate statistical analysis based on principal component analysis (16).

Secondary structure predictions

The multiple sequence alignment without the partial sequences was subjected to secondary structure prediction using the neural network-based program PHD as implemented on the World-Wide-Web (17).

General sequence analysis

The general sequence analysis tools of the GCG8 software package (18) were also used in this study.

RESULTS AND DISCUSSION

Distribution of chromo domains

During database searches for additional chromo domain sequences using a method based on profile analysis (19; see Table 1), we found that *Drosophila melanogaster* (Dm) HP1 and the related mouse proteins MOD1 and MOD2 (20) each scored highly twice, with matches at both the N- and C-termini. This internal homology can be visualised by dot matrix analysis (Fig. 1), and, although unrecognized previously in *Drosophila* HP1, was recognized by Epstein *et al.* in the two mealybug (*Planococcus citri*) HP1-related sequences (pchet1 and pchet2; 21). They attributed the repeat to a remnant of an internal gene duplication, and suggested that it may not be important for function since the C-termini of Pc, HP1, pchet1 and 2 appeared to have diverged dramatically. Further analysis (Fig. 2) revealed that the C-terminal regions of all the HP1-like proteins can indeed be aligned to the 'classical' chromo domains but form a distinct subgroup, termed here the chromo shadow domain.

Upon recognizing the distinction between the chromo and chromo shadow domains, we searched the databases with refined profiles and subsequently identified 19 new matches to the 'classical' chromo domain. They include 13 from sequences of expressed sequence tags (ESTs) or genomic sequencing projects (seven of these are shown in Figure 2 and Table 1: CeYO82, SmPAJ26, Pf0131C, CeT9A58, HuMG44, CeYK9A3, CeC29H12; and six ESTs are indicated by their EMBL database identifiers in Table 1: HSA41C081, HSAFIA009, HS10835, HS82033, HS6418, HS8101). The EST, HuMG44, is notable as it appears to be the human homologue of the *Drosophila*

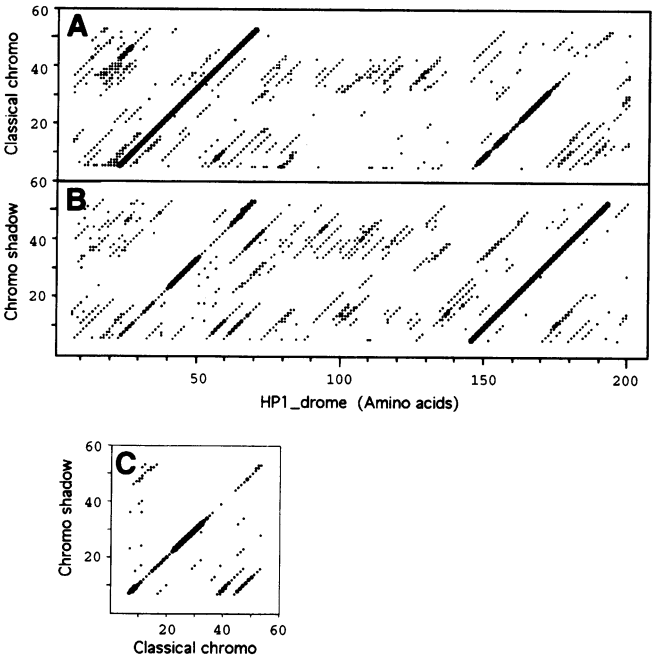

Figure 1. Dot plots, prepared with the program PROPLOT (12), of *Drosophila* HP1 amino acid sequence versus profiles based on alignments of (**A**) the classical chromo domains and (**B**) the chromo shadow domains. Values from the profile matrix which matched residues in the sequence were summed over a window of 13 residues. Large dots were plotted for the top 0.05% of the score range, medium dots for the top 0.5% and small dots for the top 10.0%. (**C**) Profile versus profile-dot plot analysis (12) using the profile based on the classical chromo domains versus that from the chromo shadow domains. Large dots were plotted for the top 0.05% of the score range, medium dots for the top 0.5% and small dots for the top 5.0%.

suppressor of PEV protein, Su(var)3-9. Both sequences also contain a SET domain (22).

MoCHD-1, HS8101 and ScYEZ4 are also potential homologues. Each possesses two chromo domains. It is possible that the *C.elegans* EST, CeYK9A3 is also a MoCHD-1 homologue, however its complete sequence is not available. MoCHD-1 was cloned by DNA affinity screening, has a novel DNA binding domain and contains a putative DEAH helicase domain (23).

Also included are four matches to sequences in putative retroelements (CfTENV, FoSKPY, MgGRH and MgMAGGY). Interestingly, the chromo domains in these sequences are all near the C-terminus of the putative polymerase polyprotein. Thus, although the data is limited, the chromo domain is arguably specific to nuclear proteins.

Structure of the chromo domains

The identification of new chromo domains permits a more accurate definition of the borders of the domain. The alignments include a chromo domain present in CeYO82. The N-terminus of the CeYO82 chomo domain is also the N-terminus of the protein itself. This indicates that the N-terminal boundary of the chromo domain has been identified. Similarly, the proximity of C-termini to the C-terminal border of both chromo (FoSKPY, MgMAGGY) and chromo shadow (DvHP1, SpSWI6) domains indicates that the C-terminal boundary has also been identified. The chromo domain is 50 aas in Pc and HP1 (aa 5–65 of Fig. 2). The chromo

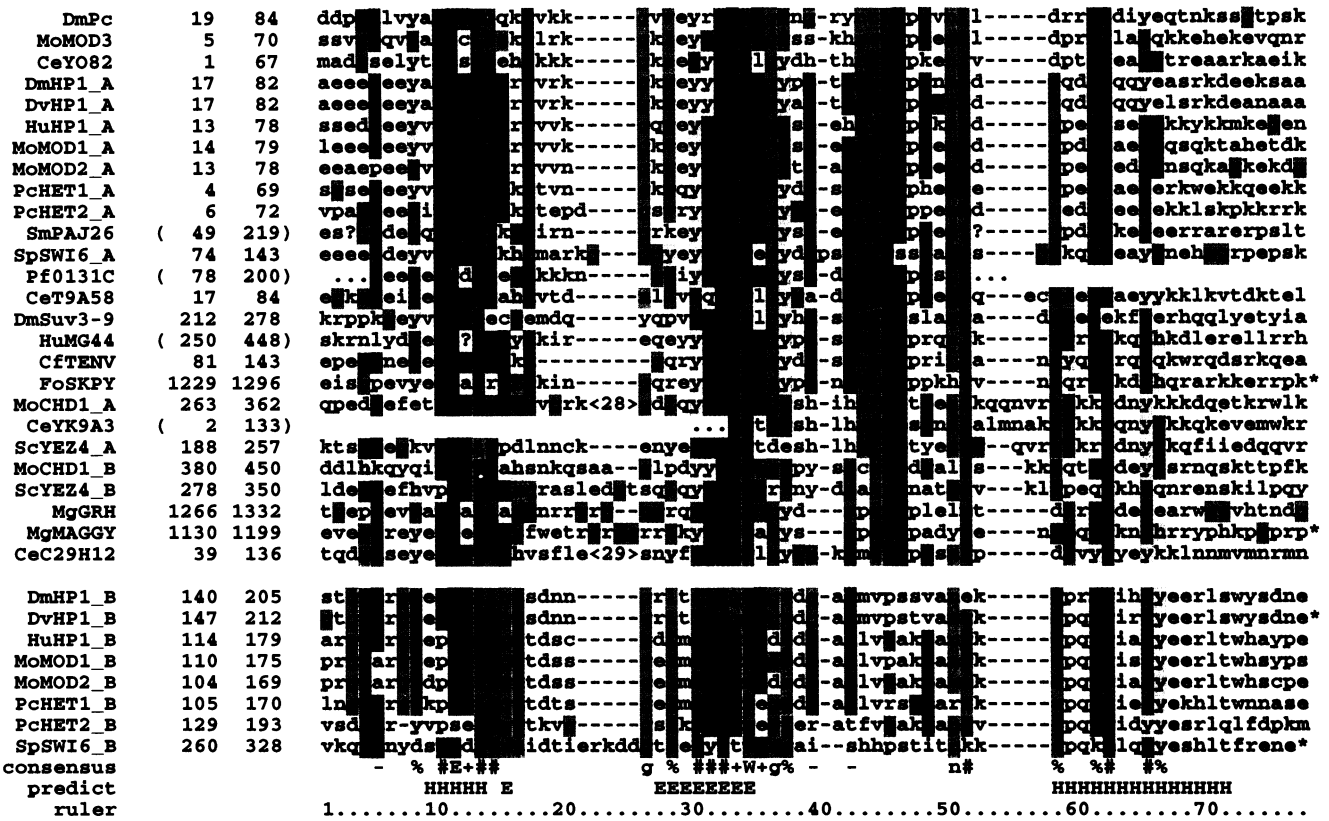

Figure 2. Alignment of chromo domains colour coded (11) according to conserved sequence similarity. The upper group of sequences contains the classical chromo domains and the lower group is the chromo shadow domains. The domain positions in the protein sequences are indicated next to their names. The numbers in parentheses refer to the positions in the DNA sequence database entries when no protein sequence is available. Frameshifted and ambiguous positions are indicated with '?' and the domains which end at the protein C-termini are ended with '*'. A consensus common to both the classical and shadow chromo domains is shown: %, semi-conserved hydrophobicity; #, strongly conserved hydrophobicity; –, conserved acidic residues; +, conserved basic residues. <28> and <29> denotes the number of residues of MoCHD1 and CeC29H12, respectively, not included in the alignment. A secondary structure prediction generated with the program PHD (17) is shown.

shadow domain is 64 aas in HP1 (aa 3–77 of Fig. 2). Although related in their N-terminal and central regions, the two chromo domains are clearly distinct in two C-terminal regions. Between positions 41 and 50 (Fig. 2) the two domains differ and the chromo shadow domain retains conservation for several residues beyond the C-terminus of the chromo domain.

From the alignment, the chromo domains can be seen as being composed of four conserved blocks. Secondary structure prediction (17) suggests a globular fold with mixed α- and β- elements. An N-terminal block that is weakly predicted to be α-helix (residues 3–17), a strongly predicted central β-strand (residues 27–38), a central block containing characteristics specific to either chromo shadow or chromo domains (residues 41–52), and a strongly predicted C-terminal α-helix (residues 59 to end). As yet, only two small mutations affecting chromo domain function have been described (24). Both are mutations of the Pc chromo domain which inactivated its ability to target to chromatin. Both (I13F; delete ID62/63) lie in the middle of the predicted α-helices and are therefore likely to perturb structure.

The chromo shadow domain

The chromo shadow domain appears in eight proteins (Fig. 2), and all available data relate these proteins to the regulation of chromatin. All eight appear to be homologues of HP1 and contain a classical chromo domain. In *Drosophila*, HP1 is found in the heterochromatin of the chromocentre of polytene chromosomes, and in a limited number of other loci, including telomeric heterochromatin (25). Its subsequent identification as the product of the *Su(var)205* locus relates its function as a chromatin regulator to its cytological characteristics (7). Functional analyses of the chromo domain of HP1 apparently conflicted with those of the Pc chromo domain. The Pc chromo domain targets Pc to specific euchromatic sites on polytene chromatin (24) whereas the HP1 chromo domain is dispensable for heterochromatin targeting of HP1 (26). In fact, the conserved C-terminal region of HP1, which we now identify as containing a chromo shadow domain, is responsible for its targeting to heterochromatin. Furthermore, a construct that does not contain the complete chromo shadow domain, does not localise to heterochromatin whereas a larger fragment which contains the complete chromo shadow domain, localises to heterochromatin (26). We therefore suggest that the chromo shadow domain is responsible for the heterochromatin targeting activity of HP1. Also notable in the chromo shadow group is the *Schizosaccharomyces pombe* SWI6 protein (27). It is involved in the repression of recombination that inappropriately activates the silent mating-type loci, *mat2* and *mat3* (28).

The chromo shadow domain occurs in proteins of similar architecture, from yeast to higher eukaryotes. All are moderately

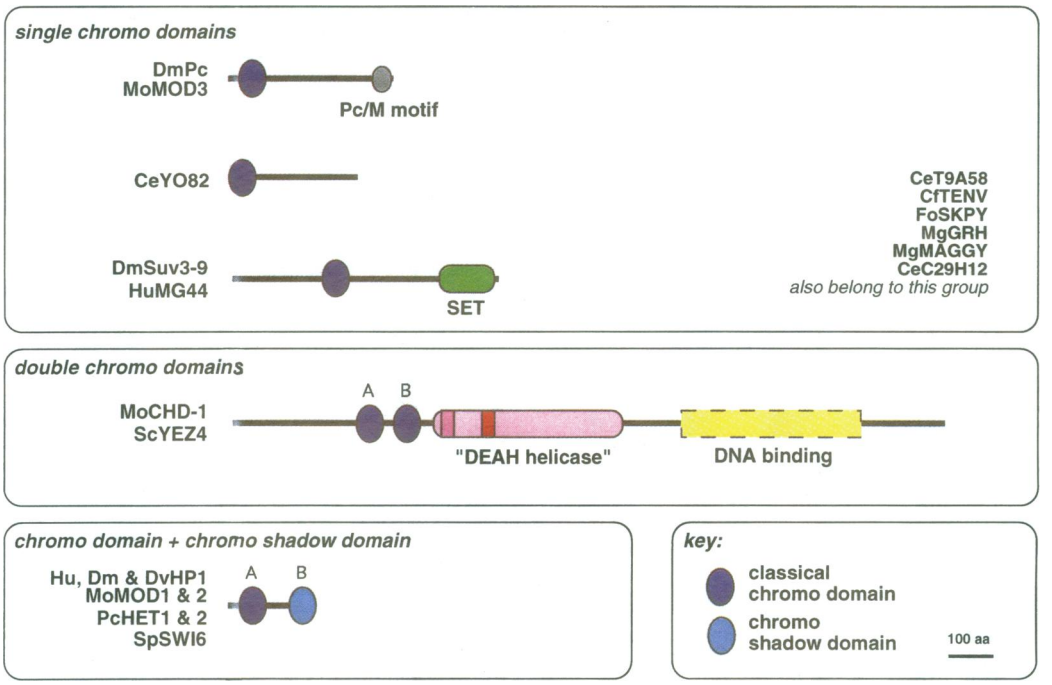

Figure 3. Organisation of chromo domain containing proteins. Chromo domains are shown as blue ovals, shadow chromo domains in light blue. Pc/M indicates a region of similarity between DmPc and MoMOD3. The SET domain is also found in the trithorax and Enhancer of zeste proteins (9,22). An uncharacterised DNA binding domain present in the MoCHD-1 protein is indicated.

sized and carry N-terminally located chromo, and C-terminally located, chromo shadow, domains (Fig. 3). This conserved disposition of chromo domains suggests that the combination invokes specific and reliable activities in chromatin regulation. It also suggests that either the two domains interact intramolecularly, or they function together as a double adaptor to bring together, by intermolecular interaction, their respective targets. Since both domains can apparently function independently to mediate specific localisation in chromatin (24,26), we favour the double adaptor model. This model envisages that chromo domains make specific interactions. If a specific chromo domain target is bound to discrete places in chromatin, then the chromo domain will bind to those places. Binding of HP1 thereby brings a second chromo domain to these sites, which would be available for further interactions. Consequently, chromatin complexes could grow, change or acquire new specificities.

Classification of chromo domains

Based on the occurrence of one or two chromo domains, the proteins identified can be divided into three categories, namely, those with a single chromo domain, those with two chromo domains and those with a chromo and a chromo shadow domain (Fig. 3). To gain further insight into the relationships between chromo domains, we performed sequence space analysis as described by Casari *et al.* (16) (Fig. 4). This method represents protein sequences, as well as sequence residues, in a multi-dimensional space which can then be projected onto a plane from any pair of dimensions. This permits the segregation of the sequences in the alignment into domain subgroups, as well as the identification of the individual residues that contribute to this segregation. Figure 4A presents a projection separating the chromo shadow domains from the classical chromo domains. The residues that plot to the extreme right (T44, W45, E46; Fig. 4B), identifies the amino acids characteristic of classical chromo domains which are not found in chromo shadow domains. Similarly V44 can be seen by this analysis as highly characteristic of chromo shadow domains.

It can also be observed that the classical chromo domains are separated, along the y-axis, by this analysis (Fig. 4A). This separation correlates to those domains that occur paired with chromo shadow domains (circled above the x-axis), and those without a paired chromo shadow (circled below the x-axis). Thus the categorisation presented in Figure 3 based simply on the mode of occurrence of chromo domains is also reflected by underlying sequence variations. As can be seen from the extremities along the y-axis (Fig. 4), the identity of amino acid 32 is the most significant residue in the y-axis spread. For chromo domains paired with chromo shadow domains, it is a leucine and for chromo domains without a paired chromo shadow domain, it is most often a valine. This correlation is conserved from yeast to higher eukaryotes suggesting that the presence of leucine or valine at position 32 is important to the combinatorial function of classical chromo domains with shadow domains.

A search for possible sequence characteristics that may underlie the third category of Figure 3, the double chromo domains (MoCHD-1 and ScYEZ4) was inconclusive. However, these chromo domains show relaxed specificity at position 32. They also show particular variance at positions 39, 41, 47 and 61 (projection data not shown), and extensions in the second indel, positions 53–57. It is possible therefore that the paired classical chromo domains define a further subset of chromo domains.

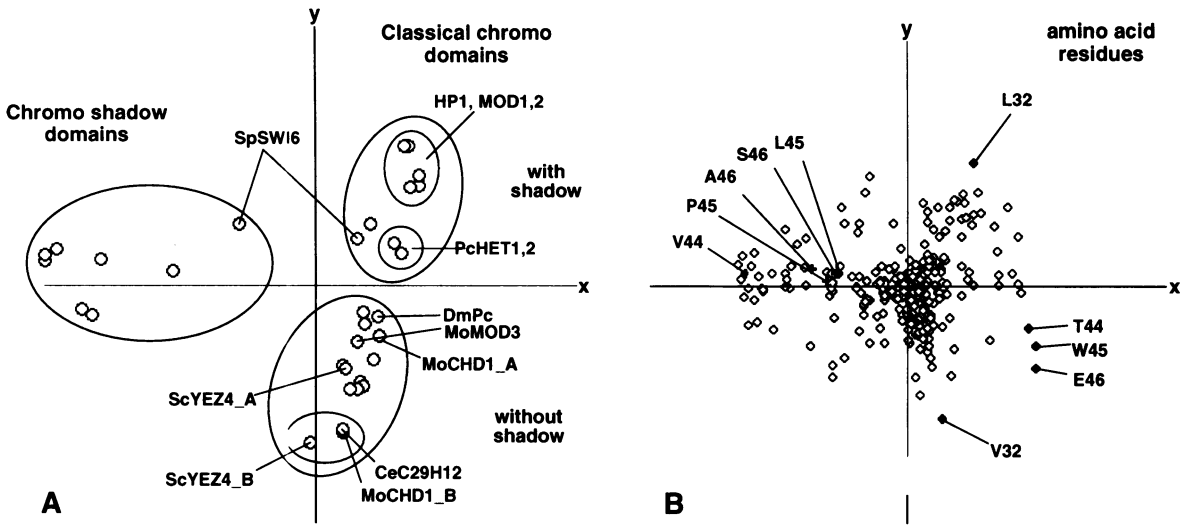

Figure 4. Sequence space analysis of the alignment of chromo domains. The projection of both plots, (A) and (B), is the same. That is, chromo shadow to the left, chromo with chromo shadow to the upper right and chromo without chromo shadow to the lower right. (**A**) The program SEQUENCE SPACE (16) was used to segregate the chromo domains into three groups: the chromo shadow domains, the classical chromo domains with, and without chromo shadow domains (indicated by the large ovals; x and y are dimensions 2 and 3, respectively). Selected examples are indicated. (**B**) The same projection as shown in (A) was used to view all the individual residues of the alignment. Selected amino acid residues are indicated. The residues which, in this projection, contribute strongly to segregation of the domain subtypes are found towards the extremities of the plot. The partial EST sequences were excluded from this analysis.

Transcriptional regulation appears to operate on at least two levels. Direct regulation is mediated by transcription factors that, upon binding DNA elements in promoters and enhancers, interact with the transcription apparatus to regulate transcriptional initiation and elongation. A second, indirect, level of regulation, appears to operate by mechanisms that either establish or maintain the repressed or activated status of chromatin regions. The apparent heterogeneity of the indirect regulators, and the difficulties inherent in studying their modes of action, present a complex puzzle. The identification of shared protein domains amongst these indirect regulators offers a basis to simplify the problem. In this paper, we have identified a new domain, the chromo shadow domain, which is apparently specific to chromatin proteins involved in indirect regulation. The chromo shadow domain is a subtype of the previously identified chromo domain. Although all known functional data relate chromo domains to the repression of chromatin, there is apparent subtype specificity in that the Pc chromo domain locates to a set of euchromatic sites in polytene spreads, whereas the HP1 chromo shadow domain appears to locate to β-heterochromatin. Our analysis also approaches the combinatorial relationships of the chromo domain subtypes and, based on sequence correlations, identifies potential key residues that may be responsible for functional specificities.

ACKNOWLEDGEMENTS

We thank Toby Gibson for his advice, encouragement and camaraderie. We thank Peter Becker and Renato Paro for discussions and Georg Casari for assistance with sequence space analysis. R.A. held fellowships from EMBO and the Norwegian Research Council.

REFERENCES

1 Kennison, J.A. and Tamkun, J.W. (1992) *New Biol.* **4**, 91–96.
2 Paro, R. (1993) *Curr. Opin. Cell Biol.* **5**, 999–1005.
3 Locke, J., Kotarski, M.A. and Tartof, K.D. (1988) *Genetics*, **120**, 181–198.
4 Orlando, V. and Paro, R. (1995) *Curr. Opin. Genet. Dev.* **5**, 174–179.
5 Paro, R. and Hogness, D.S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 263–267.
6 Paro, R. (1990) *Trends Genet.* **6**, 416–421.
7 Eissenberg, J.C., James, T.C., Foster, H.D., Hartnett, T., Ngan, V. and Elgin, S.C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9923–9927.
8 Jones, R.S. and Gelbart, W.M. (1990) *Genetics* **126**, 185–199.
9 Jones, R.S. and Gelbart, W.M. (1993) *Mol. Cell Biol.* **13**, 6357–6366.
10 Aasland, R., Gibson, T.J. and Stewart, A.F. (1995) *Trends Biochem. Sci.* **20**, 56–59.
11 Gibson, T.J., Hyvönen, M., Musacchio, A., Saraste, M. and Birney, E. (1994) *Trends Biochem. Sci.* **19**, 349–353.
12 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Comput. Applic. Biosci.* **10**, 19–29.
13 Henikoff, S. and Henikoff, J.G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
14 Clark, R.F. and Elgin, S.C. (1992) *Nucleic Acids Res.* **20**, 6067–6074.
15 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
16 Casari, G., Sander, C. and Valencia, A. (1995) *Struct. Biol.* **2**, 171–178.
17 Rost, B., Sander, C. and Schneider, R. (1994) *Comput. Applic. Biosci.* **10**, 53–60.
18 Program Manual for the Wisconsin Package, V.8., September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711.
19 Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4355–4358.
20 Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., Paro, R., James, T.C. and Gaunt, S.J. (1991) *Nucleic Acids Res.* **19**, 789–794.
21 Epstein, H., James, T.C. and Singh, P.B. (1992) *J. Cell Sci.* **101**, 463–474.
22 Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994) *EMBO J.* **13**, 3822–3831.
23 Delmas, V., Stokes, D.G. and Perry, R.P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2414–2418.
24 Messmer, S., Franke, A. and Paro, R. (1992) *Genes Dev.* **6**, 1241–1254.
25 James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A. and Elgin, S.C. (1989) *Eur. J. Cell Biol.* **50**, 170–180.
26 Powers, J.A. and Eissenberg, J.C. (1993) *J. Cell Biol.* **120**, 291–299.
27 Lorentz, A., Ostermann, K., Fleck, O. and Schmidt, H. (1994) *Gene* **143**, 139–143.
28 Klar, A.J. and Bonaduce, M.J. (1991) *Genetics* **129**, 1033–1042.